City Carrier Cost System (CCCS)
Statistical and Computer Documentation (Source Code and Data on CD-ROM)

I. PREFACE

A. Purpose and Content

USPS-FY11-34 documents the statistical design of the City Carrier Cost System (CCCS) and City Carrier Cost System – Special Purpose Route (CCCS-SPR). It contains documentation of the statistical design and the programs used to develop volume estimates for classes, products, and price categories of mail collected and delivered on city letter routes and city special purpose routes. Also included are proportions, coefficients of variation (CVs), and confidence intervals for the estimates.

B. Predecessor Documents

USPS-FY10-NP22, USPS-FY10-34, and Docket No. R2006-1, USPS-LR-L-11.

C. Corresponding Non-Public or Public Document

A nonpublic version of this document is provided as USPS-FY11-NP22.

D. Methodology

Documentation for the CCCS provided in USPS-LR-L-11, Docket No. R2006-1, included complete programs and descriptions for sample frame development and sample selection. Those programs and descriptions have incurred no substantive changes and are not reproduced herein. The estimation programs and output formats have incurred no substantive changes and are described in the CCCS System Documentation section below.

Estimates of Express Mail are now disaggregated by whether or not the carrier will attempt to obtain a signature (see Proposal Eight, Docket No. RM2011-12, approved by the PRC in Order No. 920 on October 21, 2011).

E. Input/Output

Volume estimates from the CCCS and CCCS-SPR rely on no input data. Outputs from the CCCS and CCCS-SPR are used as inputs to:

USPS-FY11-19	FY 2011 Delivery Costs By Shape
USPS-FY11-32	FY 2011 CRA "B" Workpapers (Public Version)
USPS-FY11-NP3	FY 2011 International Cost Segments and Components Report
USPS-FY11-NP14	FY 2011 CRA "B" Workpapers (Nonpublic Version)

II. ORGANIZATION

The relevant source code and outputs from the CCCS are provided on the accompanying CD-ROMs. The 'CCCS_ReadMe_FY11' file describes the contents of the CD-ROMs, which include Excel files containing proportions, coefficients of variation (CVs), and confidence intervals for both CCCS and CCCS-SPR estimates. Additionally, overviews of the statistical design and descriptions of the estimation processes are described in the System Documentation sections below.

III. CCCS SYSTEM DOCUMENTATION

A. Overview

Documentation for the CCCS provided in USPS-LR-L-11, Docket No. R2006-1, included complete programs and descriptions for sample frame development and sample selection. Those programs and descriptions have incurred no substantive changes and are not reproduced herein.

The CCCS is a continuous, ongoing cross-sectional statistical study, or probability sample of city carrier route-days. Approximately 8400 CCCS samples are scheduled each Fiscal Year. For each selected route-day, a sample of mail is selected, and for each selected mailpiece, the class, product, and other characteristics are recorded directly into a portable microcomputer using the Computerized On-Site Data Entry Systems (CODES) software.

The CCCS gathers data for distributing major portions of carriers' salaries, benefits and related costs to the categories of mail for postal rate-making and related USPS management purposes. Accrued carrier costs, available from payroll data, are total amounts and are not generally associated with any particular class of mail or service. Therefore, special methods are needed to determine the costs associated with the mail categories.

City delivery is organized and operated in terms of individual routes. Because of their different operating characteristics, routes are divided for cost development into two groups: letter routes and special purpose routes. Letter routes account for more than 95 percent of street activity costs. The CCCS considers only regular letter routes.

B. Use of CCCS Data in Cost Attribution

Total accrued labor costs for city carriers are prorated between office activity Cost Segment 6 (CS 6) and street activity Cost Segment 7 (CS 7) on the basis of time proportion estimates obtained from the In-Office Cost System (IOCS). The data from CCCS are used for apportioning street activity costs to categories of mail. Carrier street activity consists primarily of delivering mail to customers located within the zones served by city delivery. In addition it includes certain other street-related carrier activities such as delivering relays, making collections and pickups, and moving mail to and from post offices and other postal facilities.

Data from the CCCS are used to distribute volume variable costs across classes, products – including Extra Services, and price categories. The delivery portion of the

CCCS (data collected via the CODES data collection system) provides the mail category data for the distribution of volume variable mail delivery costs. The PS Form 2846 portion of the CCCS provides mail category data for the distribution of volume variable mail collection costs.

C. STATISTICAL STUDY DESIGN

The universe under study in CCCS is all mail being delivered on city letter routes. A stratified, two-stage sample design is used for CCCS. The details for each of the stages are listed below.

First Stage Sample

The first stage sample is a stratified random sample of route-days. Every city letter route is assigned to one of four strata based upon whether the route is a business or residential route, and also on the size of the route's post office (CAG A-E or F-L). Within each stratum, routes are geographically ordered, and a systematic random sample of routes is selected. Possible delivery dates (every Monday through Saturday, excluding holidays) are randomized, and systematically assigned to selected routes to determine the route-days, or first stage sample units to be enumerated. This selection process ensures both geographic and temporal dispersion of the sampled route-days, and helps control workload at the district level.

Second Stage Sample (Mailpiece)

The second stage sampling unit is a mailpiece. Parcels and accountables are usually sampled with certainty. A systematic sample of letters and flats is selected. The data collector determines the skip interval ("s") to be used – typically 10 – and the CODES software generates a random number "r", between one and "s". The data collector selects the "rth" piece, and every "sth" piece thereafter. The recommended skip interval is 10. Data collectors are allowed to change skip intervals as the need arises. The skip interval used is stored on each mailpiece record.

D. ESTIMATION AND VARIANCE

The CCCS produces two types of estimates—volumes and distribution keys (ratios). A description of the estimates is provided in the overview. Estimates are computed on a quarterly and annual basis. The annual volume estimates are the sum of the four quarterly estimates. This section provides the formulas used for FY 2011 to calculate the volumes, distribution keys, and the coefficients of variation associated with those estimates.

Notation:

y	variable of interest
w	weighting factor
h	postal quarter
i	shape domain
j	product or rate category domain
k	stratum
l	route-day
N	universe count – the number of routes in the stratum
n	completed tests in the stratum
d	delivery days in the postal quarter
s	skip utilized on a record (second stage weight)
$\hat{Y}$	estimate of the total volume
$\hat{R}$	estimate of the distribution key
Cov	estimate of the covariance
V	estimate of the variance
CV	estimate of the coefficient of variation

The weight applied to each record consists of three parts. First is the number of delivery days, d_h, in each quarter. Second is the first stage weight, indicated by N_{hk}/n_{hk}. Finally there is the skip interval, s, which is applied to each record in a test. Dividing by 1000 causes the estimates to be reported in thousands. This weighting process yields unbiased estimates of mail volumes assuming any missing tests are missed at random.

The weighting factor is:

$$w_{hk} = \left(\frac{d_h \times N_{hk} \times s}{n_{hk} \times 1000} \right)$$

Variates are defined as follows:

$$y'_{hijkl} = \begin{cases} y_{hijkl} \text{ if the unit is in the i}^{th} \textit{ and } j^{th} \text{ domains} \\ 0 \text{ otherwise} \end{cases}$$

$$x'_{hikl} = \begin{cases} x_{hikl} \text{ if the unit is in the i}^{th} \text{domain} \\ 0 \text{ otherwise} \end{cases}$$

The quarterly volume for the intersection of the i^{th} and j^{th} domains is

$$\hat{Y}_{hij} = \sum_k \sum_l w_{hk} y'_{hijkl}$$

The quarterly volume for the i^{th} domain is

$$\hat{X}_{hi} = \sum_k \sum_l w_{hk} x'_{hikl}$$

The quarterly distribution key for the intersection of the i^{th} and j^{th} domains is

$$\hat{R}_{hij} = \frac{\hat{Y}_{hij}}{\hat{X}_{hi}}$$

The annual volume for the intersection of the i^{th} and j^{th} domains is

$$\hat{Y}_{ij} = \sum_{h=1}^{4} \hat{Y}_{hij}$$

The annual volume for the i^{th} domain is

$$\hat{X}_{i} = \sum_{h=1}^{4} \hat{X}_{hi}$$

The annual distribution key for the intersection of the i^{th} and j^{th} domains is

$$\hat{R}_{ij} = \frac{\hat{Y}_{ij}}{\hat{X}_{i}}$$

Variance

In computing the sampling error on the estimates, an ultimate cluster variance estimator is used. An assumption is made that the sampling error within routes is very small relative to the overall sampling error. Therefore, the variance formula used is similar to a single-stage total or ratio estimate, except that it omits the finite population correction (fpc) factor.

The estimated stratum mean by postal quarter for the intersection of the i^{th} and j^{th} domains is

$$\bar{y}'_{hijk} = \frac{\sum_{l} y'_{hijkl}}{n_{hk}}$$

$$\hat{S}^2_{hijk} = \frac{\sum_{l} \left(y'_{hijkl} - \bar{y}'_{hijk} \right)^2}{n_{hk} - 1}$$

The estimated stratum variance for the quarterly volume for the intersection of the i[th] and j[th] domains is

$$V(\hat{Y}_{hijk}) = \frac{w_{hk}^2 \hat{S}^2_{hijk}}{n_{hk}}$$

The estimated variance for the quarterly volume for the intersection of the i[th] and j[th] domains is

$$V(\hat{Y}_{hij}) = \sum_{k} V(\hat{Y}_{hijk})$$

The estimated variance for the annual volume for the intersection of the i[th] and j[th] domains is

$$V(\hat{Y}_{ij}) = \sum_{h} V(\hat{Y}_{hij})$$

The estimated stratum mean by postal quarter for the intersection of the *i[th]* domain is

$$\bar{x}'_{hikl} = \frac{\sum_{l} x'_{hikl}}{n_{hk}}$$

$$S^2_{hik} = \frac{\sum_{l} \left(x'_{hikl} - \bar{x}'_{hik} \right)^2}{n_{hk} - 1}$$

The estimated stratum variance for the quarterly volume for the i[th] domain is

$$V(\hat{X}_{hik}) = \frac{w_{hk}^2 \hat{S}^2_{hik}}{n_{hk}}$$

The estimated variance for the quarterly volume for the i[th] domain is

$$V(\hat{X}_{hi}) = \sum_{k} V(\hat{X}_{hik})$$

The estimated variance for the annual volume for the i[th] domain is

$$V(\hat{X}_i) = \sum_h V(\hat{X}_{hi})$$

The estimated stratum covariance between the quarterly volumes for the intersection of the
i[th] and j[th] domains is

$$Cov(\hat{Y}_{hijk}, \hat{X}_{hik}) = w_{hk}^2 \hat{S}_{yxhijk}$$

where

$$\hat{S}_{yxhijk} = \frac{\sum_l \left(y'_{hijkl} - \overline{y}'_{hijk}\right)\left(x'_{hikl} - \overline{x}'_{hik}\right)}{n_{hk} - 1}$$

The estimated covariance between the quarterly volumes for the intersection of the
i[th] and j[th] domains is

$$Cov(\hat{Y}_{hij}, \hat{X}_{hi}) = \sum_k Cov(\hat{Y}_{hijk}, \hat{X}_{hik})$$

The estimated covariance between the annual volumes for the intersection of the
i[th] and j[th] domains is

$$Cov(\hat{Y}_{ij}, \hat{X}_i) = \sum_h Cov(\hat{Y}_{hij}, \hat{X}_{hi})$$

The estimated relative variance (the square of the coefficient of variation) for the quarterly distribution key for the intersection of the *i[th]* and *j[th]* domain is

$$(CV)^2(\hat{K}_{hij}) = \left(\frac{V(\hat{Y}_{hij})}{\hat{Y}_{hij}^2} + \frac{V(\hat{X}_{hi})}{\hat{X}_{hi}^2} - \frac{2Cov(\hat{Y}_{hij}, \hat{X}_{hi})}{\hat{X}_{hi}\hat{Y}_{hij}}\right)$$

The relative variance for the annual distribution key for the intersection of the *i[th]* and *j[th]* domain is

$$(CV)^2(\hat{K}_{ij}) = \left(\frac{V(\hat{Y}_{ij})}{\hat{Y}_{ij}^2} + \frac{V(\hat{X}_i)}{\hat{X}_i^2} - \frac{2Cov(\hat{Y}_{ij}, \hat{X}_i)}{\hat{X}_i\hat{Y}_{ij}}\right)$$

E. Quarterly Volume Estimates and Distribution Keys

Once the city carrier data for an entire quarter have been validated, quarterly volume estimates and distribution keys are produced. The estimated volumes are compared with the same period from the previous year and with estimates from other statistical systems. Substantial differences between the reports are investigated for additional quality assurance.

Quarterly estimation is a five-step process. First, monthly files are concatenated to form the quarterly file. Second, the weights used in the estimation procedures are produced. Third, collection mail volume estimates are calculated. Fourth, delivery volume estimates are calculated. Fifth, the Z file is produced. The quarterly estimation programs are as follows:

ALDRAN.FY2011Qq.CITY.CNTL(ALD299) is run to concatenate monthly files to form the quarterly file.

INPUTS:

Validated Monthly Data Files – DSN=ALDRAN.SHAPE.CCS11mm

Example for FY 11 month 10: ALDRAN.SHAPE.CCS1110

Only those tests that actually belong in the quarter (indicated by the first digit of the testid) are used for estimation. Below is a list of the months that should be used as inputs for the estimation for each quarter:

PQ1 includes months 10, 11, and 12.
PQ2 includes months 01, 02, and 03.
PQ3 includes months 04, 05, and 06.
PQ4 includes months 07, 08, and 09.

The program is generally run more than once, so various global analyses and edits may be performed. Additionally, weights for the second stage of sampling are applied to the data.

OUTPUTS:

The SAS dataset DSN = ALDRAN. CITY.SASDSNS.SHAPE.FILE.FY2011Qq with SAS members RAWSHP and TESTCNT. RAWSHP contains all of the raw data records for the quarter and TESTCNT includes a listing of all test identification numbers for the quarter (used for weight development).

ALDRAN.FY2011Qq.CITY.CNTL(CKEYA1) produces first-stage weights to be applied to the data received from the ALD299 program. It executes the SAS code in DSN=ALDRAN.FY2011Qq.CITY.PARMLIB(ALD750JZ) that calculates the first-stage weights applied to all weighted volume estimates.

INPUTS:

City Master frame for universe counts

DSN= ALDRAN.HQ059T01.CITYEXTR.PQ&PQ.FY&FY

Date file for number of delivery days in the quarter

DSN=ALDRAN.FY2011.PARMLIB(DATEPQq)

Data file for number of tests returned

DSN=ALDRAN.CITY.SASDSNS.SHAPE.FILE.FY2011Qq

Sample file for stratum designation

DSN=ALDRAN.PS400T01.CITY.PQqFY11

File containing validated collection tests

DSN=ALDRAN.FY11.QqDET.CCS.EDIT

Flat file containing all previously calculated weights

DSN=ALDRAN.CITY.WEIGHTS(FY2011Qq)

SAS file with weights to be used later

DSN=ALDRAN.CCS2011.PQq.YTDWGT.DATA

OUTPUTS

SAS file with weights for processing data

DSN=ALDRAN.CCS2011.PQq.YTDWGT.DATA

Flat file containing weights for processing data

DSN=ALDRAN.CITY.WEIGHTS(FY&CC&FY.Q&PQ)

ALDRAN.FY2011Qq.CITY.CNTL(CKEYA2) processes collection mail. It executes SAS code in DSN=ALDRAN.FY2011Qq.CITY.PARMLIB (ALD750X8) that calculates the weighted volumes for collection mail data.

INPUTS:

File containing validated collection tests

DSN=ALDRAN.FY11.QqDET.CCS.EDIT

SAS file with collection weights used in estimation

DSN=ALDRAN.CCS2011.PQq.YTDWGT.DATA

OUTPUTS

Quarterly collection volumes for Key Distribution

DSN=ALDRAN.LOTUS.CITY.FY2011.PQq.COLL

ALDRAN.FY2011Qq.CITY.CNTL(CKEYB1) processes delivered mail counts. It merges 1) the stratum from the sample selection file, 2) the weights for each stratum from the weights file, and 3) the mail category information from the mailcode file onto the raw mail counts file. The program then sums up the information to two levels – mailcode, for external use, and CRA Bucket, for internal use.

INPUTS

File with weights

DSN=ALDRAN.CCS2011.PQq.YTDWGT.DATA

File with mail category information for the mailcode output file

DSN=ALDRAN.FY2011Qq.SORTED.MAILCODE(CITYV5)

File with mail category information for the CRA bucket output file

DSN=ALDRAN.SASAUTOS.CTYMACRO.LIB2011(FORMATSG)

City quarterly data file (SAS file)

DSN=ALDRAN.CITY.SASDSNS.SHAPE.FILE.FY2011Qq (member

RAWSHP)

Sample file for strata

DSN=ALDRAN.PS400T01.CITY.PQqFY11

OUTPUTS

Weighted data for each mailcode (Layout 002)

DSN=ALDRAN.FY11.CITY.Qq.MCODE

Weighted data for each CRA bucket (Layout 003)

DSN=ALDRAN.FY11.CITY.Qq.CRABKT

ALDRAN.FY2011Qq.CITY.CNTL(ZFILE2) reproduces sections of the ALD299 and, CKEYA1 programs to reproduce data by testid, mailcode, and skip. The resulting SAS data file ALDRAN.CITY.Z.FY2011Qq is created for each postal quarter, converted from mainframe to PCSAS, and concatenated into one annual SAS data file. The SAS data set extension is RAWSHP.

INPUTS

City quarterly data file (SAS file)

DSN=ALDRAN.CITY.SASDSNS.SHAPE.FILE.FY2011Qq (member

RAWSHP)

File with weights

DSN=ALDRAN.CCS2011.PQq.YTDWGT.DATA

Sample file

DSN=ALDRAN.PS400T01.CITY.PQqFY11

File with mail category information

DSN=ALDRAN.FY2011Qq.SORTED.MAILCODE(CITYV5)

OUTPUTS

Quarterly Z File

DSN=ALDRAN.CITY.Z.FY2011Qq

F. Annual Estimates

Annual volume estimates are used to distribute costs to categories of mail. The volumes are calculated by summing the quarterly volumes. The annual volumes program is executed from the following location: ALDRAN.FY2011.CITY.CNTL. Two members are utilized to produce the annual volume estimates.

ALDRAN.FY2011.CITY.CNTL(SMICOLL) is used to produce annual city collection mail volumes.

INPUTS:

The quarterly volumes files:

DSN=ALDRAN.LOTUS.CITY.FY2011.PQ1.COLL

DSN=ALDRAN.LOTUS.CITY.FY2011.PQ2.COLL

DSN=ALDRAN.LOTUS.CITY.FY2011.PQ3.COLL

DSN=ALDRAN.LOTUS.CITY.FY2011.PQ4.COLL

OUTPUT:

Annual volume report files for collected mail volume estimates:

DSN= ALDRAN.LOTUS.CITY.FY2011.COLL.DATA

ALDRAN.FY2011.CITY.CNTL (SMIMCOD) is used to produce annual city delivery mail volume estimates by mailcode.

INPUTS:

The quarterly volumes files:

DSN=ALDRAN.FY11.CITY.Q1.MCODE

DSN=ALDRAN.FY11.CITY.Q2.MCODE

DSN=ALDRAN.FY11.CITY.Q3.MCODE

DSN=ALDRAN. FY11.CITY.Q4.MCODE

OUTPUT:

The annual volume file for city delivered mail volume estimates by mailcode.

DSN=ALDRAN.LOTUS.CITY.FY2011.MCODE.DATA

IV. CCCS-SPR SYSTEM DOCUMENTATION

A. Overview

Similar to the City Carrier Cost System (CCCS), the CCCS-SPR is a continuous, ongoing cross-sectional statistical study, or probability sample of SPR route-days. Approximately 1000 samples are scheduled each fiscal year. For each selected route-day, a sample of mail is selected, and for each selected mailpiece, the class, product, and other characteristics are recorded directly into a portable microcomputer using the Computerized On-Site Data Entry Systems (CODES) software.

The CCCS and CCCS-SPR gather data for distributing major portions of carriers' salaries, benefits and related costs to the categories of mail for postal rate-making and related USPS management purposes. Accrued carrier costs, available from payroll data, are aggregate amounts and are not generally associated with any particular class of mail or service. Therefore, special methods are needed to determine the costs associated with the mail categories.

City delivery is organized and operated in terms of individual routes. Because of their different operating characteristics, routes are divided for cost development into two groups: letter routes and special purpose routes. Letter routes account for approximately 95 percent of street activity costs. The CCCS considers only regular letter routes. The CCCS-SPR provides estimates of delivered mail volumes associated with the remaining street activity costs.

B. Use of CCCS-SPR Data in Cost Attribution

Total accrued labor costs for city carriers are prorated between office activity Cost Segment 6 (CS 6) and street activity Cost Segment 7 (CS 7) on the basis of time proportion estimates obtained from the In-Office Cost System (IOCS). The data from CCCS and CCCS-SPR are used for apportioning street activity costs to categories of mail. Carrier street activity consists primarily of delivering mail to customers located within the zones served by city delivery. In addition, it includes certain other street-related carrier activities such as delivering relays, making collections and pickups, and moving mail to and from post offices and other postal facilities.

Data from the CCCS-SPR are used to distribute volume variable costs across classes, products – including Extra Services, and price categories. The delivery portion of the CCCS-SPR (data collected via the CODES data collection system) provides the mail category data for the distribution of volume variable mail delivery costs for special purpose routes.

C. Statistical Study Design

The universe under study in CCCS-SPR is all mail being delivered on city special purpose routes. A stratified, two-stage sample design is used for CCCS-SPR. The details for each of the stages are listed below.

First Stage Sample

The first stage sample is a stratified random sample of route-days. Every SPR is assigned to one of four strata based upon the type of SPR operation and the number of hours clocked to street operations. Within each stratum, routes are geographically ordered, and a systematic random sample of routes is selected. Possible delivery dates (every Monday through Saturday, excluding holidays) are randomized, and systematically assigned to selected routes to determine the route-days, or first stage sample units to be enumerated. This selection process ensures both geographic and temporal dispersion of the sampled route-days, and helps control workload at the District level. Post-stratification prior to calculating the first stage weights addresses births and deaths of primary sampling units, including migrations among strata.

Second Stage Sample (Mailpiece)

The second stage sampling unit is a mailpiece. Parcels and accountables are usually sampled with certainty. A systematic sample of letters and flats is selected. The data collector determines the skip interval ("s") to be used – typically 10 – and the CODES software generates a random number "r", that can range from one to "s". The data collector selects the "r^{th}" piece, and every "s^{th}" piece thereafter. The recommended skip interval is 10. Data collectors are allowed to change the skip interval as the need arises. The skip interval used is stored on each mailpiece record.

D. Creating the Sample Frame

The sampling frame, or SPR Master Frame, is created from the most recent records from the Time and Attendance Collection System (TACS). CCCS has historically used extracts from the Address Management System (AMS) as a sampling frame for letter routes. However, AMS is not currently an option for identifying all SPRs. Therefore, an alternative method of creating a frame of SPRs was necessary. Because City Carriers must indicate a route number when clocking into LDC 23 operations, the sampling frame for the testing of SPRs is created using clock rings from recent TACS records, with the unique finance number/TACS route number being the sample unit.

Approximately five weeks prior to a new Postal Quarter (PQ), the most recent two Pay Periods of TACS SPR carrier data are extracted by executing the SAS program ALDRAN.FYyyyy.Sample.Prod.Jobs(SPRTACQq).

INPUTS

TACS file for each Area – ALB.TACS.OPRDnn.G0pppV00 where nn=area code and ppp=pay period code

OUTPUTS

TACS Extract File - ALDRAN.SPR.TACS.FYyy.PQq

Stratification and Sample Allocation

Using the TACS extract file, the SAS program ALDRAN.FYyyyy.Sample.SPR.Jobs(SPRFRMQq) is executed to develop the final sample frame, stratify the sample units, and produce sampling percentages for sample allocation.

Stratification is the process of assigning units with similar characteristics to the same group. All LDC 23 street work hours are first summed for each sample unit (finance – TACS route combination) and MODS operation number. After obtaining the total hours for each sample unit, a ratio of hours by operation number to total hours is used to classify the route type as parcel, relay, combination, or other. If the sample unit has more than 70% of the street work hours in one operation, the route type is identified by that operation. Otherwise, the route is classified as other.

Next, each sample unit is assigned either low, medium, or high based on the total hours. Due to the ad hoc nature of some LDC 23 work hour usage and the inefficiencies associated with sampling these units, sample units with less than 10 total LDC 23 street hours in the 4 week TACS data base are excluded. All others, which make up over 93% of total LDC 23 street work hours, are classified as one of three categories:

Low	10 – 39.99
Medium	40 – 199.99
High	200.00 +

Using the route type and hours usage information, each sample unit is assigned to one of four strata.

C1 Parcel, Combination, and Other routes with Medium work hour usage
C2 Parcel, Combination, and Other routes with High work hour usage
C3 Parcel and Combination routes with Low work hour usage
C4 All Relay routes and Other routes with Low work hour usage

After writing out the sample frame, the SAS program computes and saves the sampling percentages for each stratum based on the proportion of work hours in each stratum to total work hours represented by the sample frame. This proportional allocation can vary from quarter to quarter based on actual changes in work hour usage for each stratum.

INPUTS

TACS Extract File - ALDRAN.SPR.TACS.FYyy.PQq

OUTPUTS

Final Frame - ALDRAN.SPR.Frame.FYyy.PQq

Sampling Percentages – ALDRAN.SPR.SMPRCNT.FYyy.PQq

Selecting the First Stage Sample

The first stage sample is a stratified random sample of route-days. There are four steps in this process. The first step is stratification, when routes with similar characteristics are grouped. In the second step, allocation, the number of routes to be sampled in each stratum is determined. These first two steps have been included in the sample frame creation process above. In the third step, selection, routes from each stratum are randomly selected. In the fourth step, test dates are randomized and assigned to selected routes, thereby determining the route-days to be sampled. Each postal quarter (PQ), a new sample of route-days is selected, independently from those selected in prior quarters.

The SAS Program ALDRAN.FYyyyy.Sample.SPR.Jobs(SPRSMPQq) uses SAS Proc Surveyselect to produce a systematic sample of routes from each stratum. Based on sampling percentages provided in the frame creation stage, the sample size is determined for each stratum and sample routes are systematically selected. After the sample file is produced, administrative data concerning the route and finance number are added in order to merge the CCCS-SPR sample with the CCCS sample. A six-digit test identification number is assigned to each selected route-day, and is used for tracking tests throughout subsequent processing. The test identification number starts with the postal quarter number followed by a 4-digit sequential number beginning with 5001 and a one-digit check sum. The check sum is computed using the ‘MODULUS 10’ check digit algorithm.

After the CCCS-SPR sample file is produced, it is concatenated with the CCCS sample file during the CCCS and RCCS sample selection process and test dates are sequentially assigned from the randomized test date file. Test dates are assigned without replacement until all dates have been used, and then they are reused in the same randomized, sequential order.

INPUTS

Final Frame - ALDRAN.SPR.Frame.FYyy.PQq.

Sampling Percentages – ALDRAN.SPR.SMPRCNT.FYyy.PQq

OUTPUTS

Basic Sample File - ALDRAN.SPR.SMP1.FYyy.PQq

Sample File in CCCS sample format – ALDRAN.SPR.SAMPLE.FYyy.PQq

Creating the Post-Stratified Frame

At the end of each PQ, and before estimation, a post-stratified frame is created using actual PQ data from TACS. The SAS program ALDRAN.FYyyyy.Sample.SPR.Jobs(SPRPSTQq) is executed to create the post-stratified frame using the same methodology used to create the sample frame.

INPUTS

TACS PQ Extract File - ALDRAN.SPR.TACS.FYyy.PQq

OUTPUTS

Final Post-Stratified Frame - ALDRAN.SPR.POSTFRME.FYyy.PQq

Sampling Percentages – ALDRAN.SPR.SMPRCNT.FYyy.PQq

E. Estimation And Variance

The CCCS-SPR produces two types of estimates—volumes and distribution keys (ratios). A description of the estimates is provided in the overview. Volume estimates are computed on a quarterly and annual basis. The annual volume estimates are the sum of the four quarterly estimates. This section provides the formulas used for FY 2011 to calculate the volumes, distribution keys, and the coefficients of variation (CV) associated with those estimates.

Notation:

y	variable of interest
w	weighting factor
h	postal quarter
i	cost pool domain
j	product or rate category domain
k	stratum
l	route-day
T	total hours in the stratum
t	tested hours in the stratum

n	tested route days in the stratum
s	skip utilized on a record (second stage weight)
$\hat{Y}$	estimate of the total volume
$\hat{R}$	estimate of the distribution key
Cov	estimate of the covariance
V	estimate of the variance
CV	estimate of the coefficient of variation

The weight applied to each record consists of two parts. First is the first stage weight, indicated by T_{hk}/t_{hk}, which is the statum's inverse sampling fraction of the street hours, or the reciprocal of the ratio of the sampled SPR street hours to the total SPR street hours. Second is the skip interval, s, which is applied to each record in a test. Dividing by 1000 causes the estimates to be reported in thousands. .

The weighting factor is:

$$w_{hk} = \left(\frac{T_{hk} \times s}{t_{hk} \times 1000} \right)$$

Variates are defined as follows:

$$y'_{hijkl} = \left\{ \begin{matrix} y_{hijkl} \text{ if the unit is in the i}^{\text{th}} \textit{ and } j^{th} \text{ domains} \\ 0 \text{ otherwise} \end{matrix} \right\}$$

$$x'_{hikl} = \left\{ \begin{matrix} x_{hikl} \text{ if the unit is in the i}^{\text{th}} \text{domain} \\ 0 \text{ otherwise} \end{matrix} \right\}$$

The quarterly volume estimate for the intersection of the i^{th} and j^{th} domains is

$$\hat{Y}_{hij} = \sum_k w_{hk} \sum_l y'_{hijkl}$$

The quarterly volume estimate for the i^{th} domain is

$$\hat{X}_{hi} = \sum_k w_{hk} \sum_l x'_{hikl}$$

The quarterly distribution key for the intersection of the i^{th} and j^{th} domains is

$$\hat{R}_{hij} = \frac{\hat{Y}_{hij}}{\hat{X}_{hi}}$$

The annual volume for the intersection of the i^{th} and j^{th} domains is

$$\hat{Y}_{ij} = \sum_{h=1}^{4} \hat{Y}_{hij}$$

The annual volume for the i^{th} domain is

$$\hat{X}_{i} = \sum_{h=1}^{4} \hat{X}_{hi}$$

The annual distribution key for the intersection of the i^{th} and j^{th} domains is

$$\hat{R}_{ij} = \frac{\hat{Y}_{ij}}{\hat{X}_{i}}$$

Variance Estimation

In computing the sampling error on the estimates, Taylor series (first order) approximation is used. An assumption is made that the sampling error within routes is very small relative to the overall sampling error. Therefore, the variance formula used is similar to a single-stage total or ratio estimate, except that it omits the finite population correction (fpc) factor. A relative measure of sampling error, coefficient of variation (c.v.), is estimated for each annual distribution key for the intersection of the i^{th} and j^{th} domain ($\hat{R}_{ij}$).

The estimated stratum mean by postal quarter for the intersection of the i^{th} and j^{th} domains is

$$\bar{y}'_{hijk} = \frac{\sum_{l} y'_{hijkl}}{n_{hk}}$$

$$\hat{S}^2_{hijk} = \frac{\sum_l \left(y'_{hijkl} - \bar{y}'_{hijk} \right)^2}{n_{hk} - 1}$$

The estimated stratum variance for the quarterly volume for the intersection of the i[th] and j[th] domains is

$$V(\hat{Y}_{hijk}) = \frac{w^2_{hk} \hat{S}^2_{hijk}}{n_{hk}}$$

The estimated variance for the quarterly volume for the intersection of the i[th] and j[th] domains is

$$V(\hat{Y}_{hij}) = \sum_k V(\hat{Y}_{hijk})$$

The estimated variance for the annual volume for the intersection of the i[th] and j[th] domains is

$$V(\hat{Y}_{ij}) = \sum_h V(\hat{Y}_{hij})$$

The estimated stratum mean by postal quarter for the intersection of the *i[th]* domain is

$$\bar{x}'_{hikl} = \frac{\sum_l x'_{hikl}}{n_{hk}}$$

$$S^2_{hik} = \frac{\sum_l \left(x'_{hikl} - \bar{x}'_{hik} \right)^2}{n_{hk} - 1}$$

The estimated stratum variance for the quarterly volume for the i[th] domain is

$$V(\hat{X}_{hik}) = \frac{w^2_{hk} \hat{S}^2_{hik}}{n_{hk}}$$

The estimated variance for the quarterly volume for the i[th] domain is

$$V(\hat{X}_{hi}) = \sum_k V(\hat{X}_{hik})$$

The estimated variance for the annual volume for the i[th] domain is

$$V(\hat{X}_{i}) = \sum_h V(\hat{X}_{hi})$$

The estimated stratum covariance between the quarterly volumes for the intersection of the i[th] and j[th] domains is

$$Cov(\hat{Y}_{hijk}, \hat{X}_{hik}) = w_{hk}^2 \hat{S}_{yxhijk}$$

where

$$\hat{S}_{yxhijk} = \frac{\sum_l \left(y'_{hijkl} - \overline{y}'_{hijk}\right)\left(x'_{hikl} - \overline{x}'_{hik}\right)}{n_{hk} - 1}$$

The estimated covariance between the quarterly volumes for the intersection of the i[th] and j[th] domains is

$$Cov(\hat{Y}_{hij}, \hat{X}_{hi}) = \sum_k Cov(\hat{Y}_{hijk}, \hat{X}_{hik})$$

The estimated covariance between the annual volumes for the intersection of the i[th] and j[th] domains is

$$Cov(\hat{Y}_{ij}, \hat{X}_{i}) = \sum_h Cov(\hat{Y}_{hij}, \hat{X}_{hi})$$

The estimated relative variance (the square of the coefficient of variation) for the quarterly distribution key for the intersection of the *i[th]* and *j[th]* domain is

$$(CV)^2(\hat{K}_{hij}) = \left(\frac{V(\hat{Y}_{hij})}{\hat{Y}_{hij}^2} + \frac{V(\hat{X}_{hi})}{\hat{X}_{hi}^2} - \frac{2Cov(\hat{Y}_{hij}, \hat{X}_{hi})}{\hat{X}_{hi}\hat{Y}_{hij}}\right)$$

The relative variance for the annual distribution key for the intersection of the *i[th]* and *j[th]* domain is

$$(CV)^2(\hat{K}_{ij}) = \left(\frac{V(\hat{Y}_{ij})}{\hat{Y}_{ij}^2} + \frac{V(\hat{X}_{i})}{\hat{X}_{i}^2} - \frac{2Cov(\hat{Y}_{ij}, \hat{X}_{i})}{\hat{X}_{i}\hat{Y}_{ij}}\right)$$

Quarterly Volume Estimates and Distribution Keys

Once the city carrier SPR data for an entire quarter have been validated, quarterly volume estimates and distribution keys are produced. The estimated volumes are compared with the same quarter from the previous year and with estimates from other

statistical systems. Substantial differences between the reports are investigated for additional quality assurance.

Quarterly volume estimation is a four-step process. First, monthly files are concatenated to form the quarterly file. Second, the weights used in the estimation procedures are produced. Third, delivery volume estimates are calculated. Fourth, the Z file is produced. The quarterly estimation programs are as follows:

ALDRAN.FYyyyyQq.SPR.CNTL(SPR299V1) is run to concatenate monthly files to form the quarterly file. The input files are the validated monthly files.

INPUTS:

Validated Monthly Data Files – DSN=ALDRAN.SHAPE.SPRyymm

Example for FY 11 month 10: ALDRAN.SHAPE.SPR1110

Only those tests that actually belong in the quarter (indicated by the first digit of the testid) are used for estimation. Below is a list of the months that should be used as inputs for the estimation for each quarter:

PQ1 includes months 10, 11, and 12.
PQ2 includes months 01, 02, and 03.
PQ3 includes months 04, 05, and 06.
PQ4 includes months 07, 08, and 09.

The program is generally run more than once, so various global analyses and edits may be performed. Additionally, weights for the second stage of sampling are applied to the data.

OUTPUTS:

The SAS dataset DSN = ALDRAN.SPR.SASDSNS.SHAPE.FILE.FYyyyyQq with SAS members RAWSHP and TESTCNT. RAWSHP contains all of the raw data records for the quarter and TESTCNT includes a listing of all test identification numbers for the quarter.

ALDRAN.FYyyyyQq.SPR.CNTL(SKEYA1) produces first-stage weights to be applied to the data received from the SPR299V1 program. It executes the SAS code in DSN=ALDRAN.FYyyyyQq.SPR.PARMLIB(SPR750JZ) that calculates the first-stage weights applied to all weighted volume estimates.

Producing the first-stage weights is a multi-step process. First, actual TACS data for the PQ is brought in and total street hours are summarized by each finance number and route number combination for each day. Next, the TACS summary data is merged with the quarterly route file and the actual street hours tested (n) is determined for each stratum. Next, the TACS summary data is merged with the post-stratified frame to determine the total number of actual street hours (N) used for each stratum. Finally, weights are created for each stratum and saved for use in later estimation processes.

INPUTS:

TACS Pay Period files for actual street hours summaries

DSN=ALDRAN.SPR.TACS.FYyy.PPpp

SPR Post-stratified frame for universe summary

DSN= ALDRAN.SPR.POSTFRME.FYyy.PQq

Monthly Route Files for tested route summaries

DSN= ALDRAN.ROUTE.SPRyymm

Quarterly Shape File for analysis

DSN=ALDRAN.SPR.SASDSNS.SHAPE.FILE.FYyyyyQq

Sample files for stratum designation

DSN=ALDRAN.SPR.SAMPLE.FYyy.PQq

DSN= ALDRAN.SPR.SMP1.FYyy.PQq

OUTPUTS

Flat file with weights for processing data

DSN=ALDRAN.SPR.WEIGHTS(FYyyyyQq)

SAS file with weights for processing data

DSN=ALDRAN.SPRyyyy.PQq.YTDWGT.DATA

Quarterly Route file containing actual tested street hours for analysis

DSN=ALDRAN.SPR.ROUTE.FYyyyyQq

ALDRAN.FYyyyyQq.SPR.CNTL(SKEYB1) processes delivered mail counts. It merges 1) the stratum from the sample selection file, 2) the weights for each stratum from the weights file, and 3) the mail category information from the mailcode file onto the raw mail counts file. The program then sums up the information to two levels – mailcode, for external use, and CRA Bucket, for internal use.

INPUTS

File with weights

DSN=ALDRAN.SPRyyyy.PQq.YTDWGT.DATA

File with mail category information for the mailcode output file

DSN=ALDRAN.FYyyyyQq.SORTED.MAILCODE(CITYV5)

File with mail category information for the CRA bucket output file

DSN=ALDRAN.SASAUTOS.CTYMACRO.LIByyyy(FORMATSG)

City quarterly data file (SAS file)

DSN=ALDRAN.SPR.SASDSNS.SHAPE.FILE.FYyyyyQq (member

RAWSHP)

Sample file for strata

DSN=ALDRAN.PS400T01.CITY.PQqFYyy

OUTPUTS

Weighted data for each mailcode

DSN=ALDRAN.FYyy.SPR.Qq.MCODE

Weighted data for each CRA bucket

DSN=ALDRAN.FYyy.SPR.Qq.CRABKT

ALDRAN.FYyyyyQq.SPR.CNTL(ZFILE3) reproduces sections of the SPR299V1 and, SKEYA1 programs to reproduce data by testid, mailcode, and skip. The resulting SAS data file ALDRAN.SPR.Z.FYyyyyQq is created for each postal quarter, converted from mainframe to PCSAS, and concatenated into one annual SAS data file. The SAS data set extension is RAWSHP.

INPUTS

City quarterly data file (SAS file)

DSN=ALDRAN.SPR.SASDSNS.SHAPE.FILE.FYyyyyQq (member

RAWSHP)

File with weights

DSN=ALDRAN.SPRyyyy.PQq.YTDWGT.DATA

Sample file

DSN=ALDRAN.PS400T01.CITY.PQqFYyy

File with mail category information

DSN=ALDRAN.FYyyyyQq.SORTED.MAILCODE(CITYV5)

OUTPUTS

Quarterly Z File

DSN=ALDRAN.SPR.Z.FYyyyyQq

Annual Estimates

Annual volume estimates are used to distribute costs to categories of mail. First, the quarterly Z files are concatenated to form an annual Z file. Next, the volumes for the applicable categories are calculated by running the program SPR_Output_V7.sas.

INPUT:

The annual Z file with all parcel and accountable entries:

DSN=SPR_Z_ACRFY11_final

OUTPUT:

The annual keys file for SPR delivered parcels and accountables.

SPR_Output_FY11_Final.xls

City Z File Layout - 001

The variable names and explanations follow.

SAS Variable Name	Description
BKTCHAR	Letter Character
BKTNUM	Bucket Number
COMPLETE	Total number of completed delivery tests in the quarter
DELDAYS	Delivery days in the quarter
DELWGT	The first stage weight
F2846	Total number of collection mail forms completed in the quarter
F28WGT	Weight assigned to collection mail strata
MAILCODE	Mailcode for the record
MASTER	Stratum universe count of routes
NAME	Description of mailcode
SKIP	Skip interval for record (second stage weight)
STRATUM	Stratum in which the route (testid) exists
NOPIECES	Total mailpieces for the entry weighted by the skip interval
TESTID	Identification number for test
WGT	DELWGT/1000

City Mcode File – Layout 002

Position	Description
1 - 15	Volume
18 - 23	Mailcode
25 - 27	Bucket number
30 - 80	Mailcode description

Bucket Descriptions Layout - 003

Bucket	Description
001	'FIRST-CLASS MAIL
111	' SINGLE PIECE LETTERS
112	' SINGLE PIECE FLATS
113	' SINGLE PIECE PARCELS
121	' PRESORT LETTERS
122	' PRESORT FLATS
123	' PRESORT PARCELS
141	' SINGLE PIECE CARDS
151	' PRESORT CARDS
189	' TOTAL FIRST-CLASS MAIL
190	'PRIORITY MAIL
200	'EXPRESS MAIL-WITH SIGNATURE
201	'EXPRESS MAIL-NO SIGNATURE
210	'PERIODICALS
300	'STANDARD MAIL-REG AND NONPROFIT
311	' STANDARD OTHER LETTERS
312	' STANDARD OTHER FLATS
313	' STANDARD OTHER PARCELS
320	' TOTAL STANDARD OTHER
330	'
331	' ECR BASIC LETTERS
332	' ECR BASIC FLATS
333	' ECR BASIC PARCELS
350	'
351	' ECR HI-DENSITY LETTERS
352	' ECR HI-DENSITY FLATS
353	' ECR HI-DENSITY PARCELS
360	'
361	' ECR SATURATION LETTERS
362	' ECR SATURATION FLATS
363	' ECR SATURATION PARCELS
370	' TOTAL ECR
380	'
390	' NOT FLAT MACHINABLE (NFM)
395	'
399	' TOTAL STANDARD
400	'
401	'PACKAGE SERVICES
410	' PARCEL POST SINGLE PIECE
420	' PARCEL SELECT
430	'
442	' BOUND PRINTED MATTER FLATS

Bucket Descriptions Layout - 003

443	' BOUND PRINTED MATTER PARCELS
450	' MEDIA AND LIBRARY
460	'
490	' TOTAL PACKAGE SERVICES
600	'
610	'U.S. POSTAL SERVICE
620	'FREE MAIL - - BLIND & HNDC
630	'
700	'INTERNATIONAL MAIL
710	' LETTERS/CARDS/AO
720	' PARCEL POST
730	' CANADA EXPEDITED PARCEL
740	' CANADA XPRESSPOST
800	’
810	'
820	'
830	'
840	' INTERNATIONAL EXPRESS
850	'
880	'TOTAL INTERNATIONAL MAIL
890	'TOTAL DOMESTIC MAIL
900	'TOTAL ALL MAIL
901	' ACCT POSTAGE DUE
902	' ACCT BUSINESS REPLY
903	' ACCT CERTIFIED
904	' ACCT COD
905	' ACCT NUMBERED INSURED
906	' ACCT REGISTERED
907	' ACCT RETURN RECEIPT
908	' ACCT DELIVERY CONFIRMATION
909	' ACCT SIGNATURE CONFIRMATION
910	' ACCT OTHER
920	' USPS PFS
990	'OTHER MAIL CLASS
999	‘COMPETITIVE PRODUCTS

City Collection File Layout - 004

1 - 2 Line Number
4 - 23 Rate Category
26 – 36 Customer Outgoing Letter and Flat Volumes
38 – 48 Customer Outgoing Parcel Volumes
50 – 60 Collection Box Letter and Flat Volumes
62 – 72 Collection Box Parcel Volumes
74 – 84 Carrier Pickup
86 – 96 Customer Outgoing Total
98 – 108 Collection Box Total

SPR Z File Layout - 005

The variable names and explanations follow.

SAS Variable Name	Description
BKTCHAR	Letter Character
BKTNUM	Bucket Number
COUNT	Total mailpieces for the entry.
DELWGT	The first stage weight
MAILCODE	Mailcode for the record
NAME	Description of mailcode
NOPIECES	Total mailpieces weighted by the skip interval
RTDAY_HRS	LDC 23 street hours used on the test day
SAMP_STRATA_HRS	Total sampled LDC 23 street hours for the stratum
SKIP	Skip interval for record (second stage weight)
SS1	First Extra Service for the entry
SS2	Second Extra Service for the entry
SS3	Third Extra Service for the entry
SS4	Fourth Extra Service for the entry
STRATA	Stratum in which the sample unit exists
STRATA_HRS	Total LDC 23 street hours for the stratum
TESTID	Identification number for test
WGT	DELWGT/1000